Regularized estimation of the nominal response model

Abstract

The nominal response model is an item response theory model that does not require the ordering of the response options. However, while providing a very flexible modeling approach of polytomous responses, it involves the estimation of many parameters at the risk of numerical instability and overfitting. The lasso is a technique widely used to achieve model selection and regularization. In this paper, we propose the use of a fused lasso penalty to group response categories and perform regularization of the unidimensional and multidimensional nominal response models. The good performance of the method is illustrated through real-data applications and simulation studies.

Keywords: Adaptive Lasso, Collapse; Fused Lasso; Item Response Theory; Lasso; Multidimensional; Polytomous Responses

1 Introduction

Many tests and questionnaires include questions with polytomous responses. Within Item Response Theory (IRT), these can be modeled if regarded as manifestations of some underlying latent variables (for a broad review on unidimensional models see for example Bartolucci et al. (2015), while multidimensional models are extensively treated in Reckase (2009)). Some models, such as the graded response model (Samejima, 1969) and the generalized partial credit models (Muraki, 1992), assume that the response categories have a predetermined order. Instead, the nominal response model (Bock, 1972) does not require an ordering of the response categories. It is used i) to model purely nominal response categories, ii) to verify empirically the expected order of the response categories, and iii) to model the responses of sets of items called testlets (Thissen et al., 2010). It is certainly the most flexible parametric IRT model for polytomous responses, and includes the generalized partial credit model and the 2-parameter logistic (2PL) model as special cases (Thissen & Steinberg, 1986). On the other hand, it conflicts with the general principle of parsimony, as it involves many parameters, which can result in numerical instability and overfitting, especially in small samples. In this respect, Thissen & Steinberg (1986) proposed an interesting parameterization of the model, which is able not only to represent the nominal model and the more constrained generalized partial credit model, but also models between these two extremes. This is achieved using polynomial bases (Thissen & Steinberg, 1986) or Fourier basis (Thissen et al., 2010), and provides smooth parameter estimates.

In recent years, penalty-based methods have attracted increasing interest in the statistical community because of their effectiveness in performing regularization and prevent overfitting (Hastie et al., 2009, 2015). The penalty term constrains the parameters of the model and, in some cases, leads to a more parsimonious one where some coefficients are set to zero or are forced to be equal. Particularly suited for this task (Hastie et al., 2015) are lasso-type penalties, which were first proposed by Tibshirani (1996) for the linear regression model to shrink the coefficients toward zero. Important extensions of this method include the fused lasso (Tibshirani et al., 2005), which forces the coefficients to assume the same value, and the group lasso (Yuan & Lin, 2006), which constrains sets of coefficients toward zero.

Originally, penalty-based methods were proposed for metrical variables. However, in recent years, regularization methods have been used for categorical data as well. Tutz & Gertheiss (2016) give a broad review of regularization methods used for both predictor and response categorical variables. Of particular interest here are the multinomial logit models (Tutz et al., 2015), where a nominal response variable is related to some manifest predictors. Tutz & Schauberger (2015) proposed the use of lasso-type penalties in the Rasch model for the detection of differential item functioning, while a different approach for the detection of differential item functioning is proposed in Magis et al. (2015), where a lasso penalty is used in a logistic regression model.

In this paper the use of a fused lasso approach for the nominal response model is explored, considering both the unidimensional and the multidimensional cases. The objectives of the proposal are multifold. First, the method is aimed at grouping the response categories, thus leading to a smaller number of parameters to estimate. Furthermore, it is intended to regularize the estimates and hence reduce the bias and the variability. The procedure is also able to set some parameters exactly to zero, when this is supported by the data. In this case, the item is no longer related to the latent variable, hence performing also variable selection. This can be of particular interest in multidimensional models, to understand which latent variables are actually related to the items and to provide sparse solutions.

2 Preliminaries

2.1 The nominal response model

Let Y_{ij} be the response of subject i to item j. In the nominal response model, the probability of giving response k is given by

$$P(Y_{ij} = k|\theta_i) = \frac{e^{\alpha_{jk}\theta_i + \beta_{jk}}}{\sum_{h=0}^{m_j-1} e^{\alpha_{jh}\theta_i + \beta_{jh}}}, \quad k = 0, \dots, m_j - 1, \tag{1}$$

where m_j is the number of response categories of item j, θ is a latent variable, α_{jk} are slope parameters that quantify the relation between the item's responses and the latent variable, and β_{jk} are intercepts (Thissen & Cai, 2016). A constraint on the parameters is needed to ensure identifiability, and in this paper α_{j0} and β_{j0} are set to 0 $\forall j$. When $m_j = 2$, Equation (1) gives the response probabilities of the 2PL model (Birnbaum, 1968). The generalized partial credit model can also be obtained as a constrained version of the nominal model (Thissen et al., 2010).

The method commonly used to estimate the parameters of the nominal model is the marginal likelihood method (Bock & Aitkin, 1981), which requires the maximization of the marginal log-likelihood function

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \log \int_{\mathbb{R}} \prod_{j=1}^{J} \prod_{k=0}^{m_j-1} P(Y_{ij} = k|\theta_i)^{I(Y_{ij}=k)} \phi(\theta_i) d\theta_i, \tag{2}$$

where J is the number of items, n is the number of subjects, $\boldsymbol{\alpha}$ is a vector containing all the slope parameters, $\boldsymbol{\beta}$ is a vector containing all the intercept parameters, $I(\cdot)$ denotes the indicator function, which is equal to 1 if its argument is true and to 0 otherwise, and $\phi(\cdot)$ denotes the density of the standard normal variable. The integral in Equation (2) does not have a closed-form solution and it is usually approximated using Gaussian quadrature.

The multidimensional nominal response model relates the probability of giving response k to more than one latent variable and requires a different slope parameter for each category and each dimension, thus leading to a very large number of parameters (Thissen & Cai, 2016). Denoting by $\boldsymbol{\theta}_i$ the D-dimensional vector of latent variables of subject i, the model can be written as:

$$P(Y_{ij} = k|\boldsymbol{\theta}_i) = \frac{e^{\sum_{d=1}^{D} \alpha_{jdk}\theta_{id} + \beta_{jk}}}{\sum_{h=0}^{m_j-1} e^{\sum_{d=1}^{D} \alpha_{jdh}\theta_{id} + \beta_{jh}}}, \quad k = 0, \ldots, m_j - 1, \tag{3}$$

with $\alpha_{jd0} = 0 \ \forall j, d$ for identifiability. The log-likelihood function is then:

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \log \int_{\mathbb{R}^D} \prod_{j=1}^{J} \prod_{k=0}^{m_j-1} P(Y_{ij} = k|\boldsymbol{\theta}_i)^{I(Y_{ij}=k)} \phi(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i, \tag{4}$$

where $\phi(\cdot)$ denotes the density of the multivariate standard normal.

2.2 The lasso method

The lasso (least absolute shrinkage and selection operator) was first proposed for the linear regression model as a method to improve prediction accuracy and interpretation (Tibshirani, 1996). The method consists in adding a constraint to the least square problem with the effect of shrinking some coefficients and setting others to zero. When the lasso problem is written in Lagrangian form, it corresponds to the minimization of a loss function with a penalty term (Hastie et al., 2015). This approach can be extended naturally to likelihood-based models (Fan & Li, 2001). Let $\ell(\boldsymbol{\theta})$ be the log-likelihood function of a parameter vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_J)^\top$, the maximization of

$$\ell_p(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - J_\lambda(\boldsymbol{\theta}) \tag{5}$$

leads to a penalized likelihood estimator, where $J_\lambda(\boldsymbol{\theta})$ is the penalty term. In the original lasso method the penalty term is the L_1-norm multiplied by the tuning parameter λ:

$$J_\lambda(\boldsymbol{\theta}) = \lambda ||\boldsymbol{\theta}||_1 = \lambda \sum_{j=1}^{J} |\theta_j|. \tag{6}$$

The fused lasso (Tibshirani et al., 2005) is a generalization of the lasso method that encourages sparsity of the differences of the coefficients besides of sparsity of the coefficients. Assuming that the coefficients θ_j have a natural order, the fused lasso penalty is given by

$$J_\lambda(\boldsymbol{\theta}) = \lambda_1 \sum_{j=1}^{J} |\theta_j| + \lambda_2 \sum_{j=2}^{J} |\theta_j - \theta_{j-1}|. \tag{7}$$

3 A proposal for the nominal response model

In the nominal response model, when the slope parameters of different categories of the same item are equal, these categories can be collapsed (Thissen & Cai, 2016). In fact, they have the same relation to the latent variable, except for a different probability of being selected that is expressed by the intercepts (Thissen et al., 2010). For this reason, the penalty term proposed in this paper encourages the slope parameters of the same item to assume the same value. Considering the unidimensional model, the penalized log-likelihood function is as follows:

$$\ell_p(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \lambda \sum_{j=1}^{J} \sum_{k=0}^{m_j-2} \sum_{h=k+1}^{m_j-1} |\alpha_{jk} - \alpha_{jh}|. \tag{8}$$

The penalization in Equation (8) is similar to the fused lasso penalty (7). However, in this context, there is not a natural order of the slope coefficients and hence all pairs of coefficients of the same item should be considered. Since $\alpha_{j0} = 0 \ \forall j$, the penalty constrains the slope parameters toward zero without requiring another penalty term. In the extreme case that all the slope parameters of one item are zero, the latent variable is no longer related to the probability of selecting the response categories. In this case, these probabilities are described only by the intercepts and depend on the frequencies. Thus, the method allows also to select the items that are actually related to the latent variable.

A limitation of the lasso is the inconsistency of the estimator under certain conditions (Zou, 2006). To overcome this drawback, Zou (2006) proposed an adaptive version of the method that includes data-dependent weights in the penalty. In this paper, we explored an adaptive version of the penalty, which requires the maximization of the following function:

$$\ell_p(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \lambda \sum_{j=1}^{J} \sum_{k=0}^{m_j-2} \sum_{h=k+1}^{m_j-1} |\alpha_{jk} - \alpha_{jh}| \, w_{jkh}, \tag{9}$$

with

$$w_{jkh} = |\tilde{\alpha}_{jk} - \tilde{\alpha}_{jh}|^{-1}, \tag{10}$$

where $\tilde{\alpha}_{jk}$ denotes the estimates obtained applying a small lasso penalization with λ fixed at 0.001. We chose not to use the maximum likelihood estimates because of their instability in small samples. Since the tuning parameter is fixed at a small value, the consistency of the estimator is preserved as the effect of this penalty is negligible in large samples. A similar solution was adopted also by Masarotto & Varin (2012), which used a small ridge penalty.

In the multidimensional case, the slopes related to each item and each dimension are forced to assume a common value. This is achieved considering the following penalized log-likelihood function:

$$\ell_p(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \lambda \sum_{j=1}^{J} \sum_{d=1}^{D} \sum_{k=0}^{m_j-2} \sum_{h=k+1}^{m_j-1} |\alpha_{jdk} - \alpha_{jdh}|. \tag{11}$$

As a result, the slopes of the same item can be grouped in a different way across the various dimensions. If all the slope parameters related to one dimension are set to zero, that item is no longer related to that dimension. So, the method is also suitable to explore which dimensions contribute to each item. The adaptive version includes the weights $w_{jdkh} = |\tilde{\alpha}_{jdk} - \tilde{\alpha}_{jdh}|^{-1}$ in the penalized log-likelihood function:

$$\ell_p(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \lambda \sum_{j=1}^{J} \sum_{d=1}^{D} \sum_{k=0}^{m_j-2} \sum_{h=k+1}^{m_j-1} |\alpha_{jdk} - \alpha_{jdh}| \, w_{jdkh}. \tag{12}$$

Multidimensional latent variable models usually require a constrain on the slope parameters to assure a unique solution to the maximization problem because of the invariance under rotation (Reckase, 2009, Chapter 6). In fact, the log-likelihood function returns the same value when the slope parameters are rotated. However, the penalty term considered here is not invariant under rotation, thus assuring a unique solution without requiring any additional constrain.

3.1 Implementation

All analyses were performed in R (R Core Team, 2018). Parts of code were developed in C++ language to speed up the computational time, using the packages `Rcpp` (Eddelbuettel & François, 2011) and `RcppArmadillo` (Eddelbuettel & Sanderson, 2014). In order to estimate the parameters of the model, we explored various algorithms. In view of the similarity of the penalty employed in this paper and the penalty used in Masarotto & Varin (2012) to rank sport teams fitting a penalized Bradley-Terry model, we first implemented the algorithm adopted in that paper. This algorithm was also proposed in Wang et al.

(2013) for the fused lasso signal approximator problem, and uses the alternating direction method of multipliers algorithm (Hastie et al., 2015, Section 5.7). Another method we implemented is proximal gradient (Hastie et al., 2015, Section 5.3.3), using the R package genlasso (Arnold & Tibshirani, 2014) for the computation of the proximal operator. Finally, we employed a quadratic approximation of the absolute value function as suggested in Tutz & Gertheiss (2014). Specifically, the absolute value of $\alpha_{jk} - \alpha_{jh}$, which enters in equations (8) and (9), was approximated with the function $\sqrt{(\alpha_{jk} - \alpha_{jh})^2 + c}$, where c is a very small value. The absolute value in equations (11) and (12) was approximated in the same way. Differently from the absolute value, this function is differentiable and thus allows to use the optimization algorithms available in standard software. When $c = 0$ the function returns the absolute value, hence the approximation is more accurate for smaller values of c. In our implementation, the penalized log-likelihood function was maximized using the BFGS optimization algorithm implemented in the R function `optim` for decreasing values of c, using the previous solution as initial values. Starting from $c = 0.01$, it was updated using the rule $c^{t+1} = c^t \cdot 10^{-1}$, until $c = 10^{-10}$. In order to improve the accuracy of the solution found by the optimization procedure, the gradient of the penalized log-likelihood function was also given as input. We first computed the maximum likelihood estimates, which were obtained setting $\lambda = 0$ in the log-likelihood function and using the estimates obtained with the `mirt` package (Chalmers, 2012) as initial values. The estimates obtained with `mirt` were not considered as the MLE estimates for two reasons: the maximization of the log-likelihood function was able to reach a higher value, and furthermore the parameterization of the nominal model in the multidimensional case used in `mirt` is a constrained version of the model as given in Thissen et al. (2010). Using the MLE as initial values, the small lasso penalization estimates $\tilde{\alpha}_{jk}$ were then obtained. The procedure was then repeated for increasing values of λ, using the previous estimates as initial values. The integrals in Equations (2) and (4) were approximated using the Gauss-Hermite quadrature with 61 points in the unidimensional case, and 31 points in the multidimensional case. In our experiments, the quadratic approximation gave better results, since the penalized log-likelihood function evaluated in the solution found by the algorithm was equal or higher than the values achieved with the former two methods. Hence, the quadratic approximation was the method adopted in the application and the simulation studies presented in the subsequent sections. The R package `regIRT` implements the methods and is provided as supplemental material.

4 Real-data examples

4.1 Unidimensional case

The application of the proposal in the unidimensional case is illustrated using data from the 2016 European Social Survey, which are available at https://www.europeansocialsurvey.org/data. In particular, the following questions on climate change were considered in this analysis:

- D2 In your daily life, how often do you do things to reduce your energy use? This item is scored as 1=never, 2=hardly ever, 3=sometimes, 4=often, 5=very often, 6=always, 55=cannot reduce energy use. [rdcenr]

- D23 To what extent do you feel a personal responsibility to try to reduce climate change? This item is scored from 0 to 10, where 0=not at all and 10=a great deal. [ccrdprs]

- D24 How worried are you about climate change? This item is scored as 1=not at all worried, 2=not very worried, 3=somewhat worried, 4=very worried, 5=extremely worried. [wrclmch]

- D25 How good or bad do you think the impact of climate change will be on people across the world? This item is scored from 0 to 10, where 0=extremely bad and 10=extremely good. [ccgdbd]

For our analyses, we selected the 1551 respondents from Sweden. Only one person gave response "cannot reduce energy use" in item D2, so it was treated as a missing response.

Two records that have completely missing responses in these four items were deleted from the dataset, thus leading to a sample size of 1549. Figure 1 shows the regularization path of two items chosen as example, while Table 1 reports the estimates of the slope parameters at the selected value of λ. The first category is not shown because the coefficients are fixed at 0. The maximum likelihood estimates of the slope parameters of item *rdcenr* are ordered with quite close values. The non-adaptive penalization leads to a fusion of the first and the third categories, while the second one was not at the selected value of λ. Instead, the adaptive penalty keeps separate all the categories. Item *ccrdprs* has 11 response categories with maximum likelihood estimates not ordered. The non-adaptive penalization groups some of the slope parameters, however they are not ordered. The adaptive penalization results in a more reasonable solution, leading to a fusion of the categories from 2 to 6 and returning ordered slopes. It is worth noting that, despite the penalty involves only the slope parameters, the intercept parameters were regularized as well.

[Figure 1 about here.]

[Table 1 about here.]

4.2 Multidimensional case

The regularized estimation of the multidimensional nominal model is illustrated using data from the Consumer Protection and Perceptions of Science and Technology section of the 1992 Eurobarometer Survey, based on a sample from Great Britain. These data were analyzed also in Bartholomew et al. (2008, Chap. 9), and can be downloaded from the book's website. The questions, which are rated on a four-point scale, are given below:

1. Science and technology are making our lives healthier, easier and more comfortable. [*Comfort*]

2. Scientific and technological research cannot play an important role in protecting the environment and repairing it. [*Environment*]

3. The application of science and new technology will make work more interesting. [*Work*]
4. Thanks to science and technology, there will be more opportunities for the future generations. [*Future*]
5. New technology does not depend on basic scientific research. [Technology]
6. Scientific and technological research do not play an important role in industrial development. [*Industry*]
7. The benefits of science are greater than any harmful effects it may have. [*Benefit*]

Consistently with the findings of Bartholomew et al. (2008), items *Comfort*, *Work*, *Future* and *Benefit* are mainly related to one dimension, while items *Environment*, *Technology* and *Industry* are mainly related to another dimension. This could be partially explained explained by the negative connotation of items *Environment*, *Technology* and *Industry.* Figure 2 shows the regularization path of two items chosen as example. Item *Environment* has negative slopes in the first dimension, which tend to be collapsed using both the non-adaptive and the adaptive penalization. Regarding the second dimension, this item has one highly positive slope and two negative slopes. Both types of penalization shrink the negative values toward zero, thus suggesting that this behavior is due to random variability. Instead, item *Work* presents positive slopes in the first dimension and negative slopes in the second. The negative slopes shrink toward zero using both the penalizations. We can also observe that the adaptive version is more effective in leading to a fusion of the categories. Table 2 reports the estimates obtained with the adaptive penalization at the selected value of λ, which are then used as true values in the simulation study.

[Figure 2 about here.]

[Table 2 about here.]

5 Simulation studies

5.1 Unidimensional case

Some simulation studies examined the performance of the proposal. The first simulation study explored the unidimensional model. The true item parameters were set equal to values used in Wollack et al. (2002), considering only the first four items. They were reparameterized so that the values of the first category were set to zero, as reported in Table 3. The slope parameters α_{12} and α_{13} were set to the same value on purpose to observe whether the procedure proposed in this paper indicates that they should be fused. Figure 3 represents the probability curves of these items.

[Table 3 about here.]

[Figure 3 about here.]

The values of the latent variable θ were generated from a standard normal distribution. The sample size was taken equal to $n = 200$, 500, 1000 and 5000. The values taken for the regularization parameter λ depend on the sample size, because in larger samples a larger value of λ is necessary to achieve the same amount of shrinkage. More specifically, λ was taken equal to 50 equispaced values from 0 to 5, 10, 15, or 20. The value of λ was then selected using 5-fold cross-validation and the BIC. All results were based on 500 replications for each setting.

A relevant result to observe is whether the procedure indicates that the slope parameters α_{12} and α_{13} should be fused. Table 4 reports the number of cases out of 500 in which these parameters are fused at the selected value of λ. These results show that the adaptive penalization performs better than the non-adaptive penalization, and that cross-validation is more effective than BIC in selecting the tuning parameter λ. It appears BIC gives good results for $n = 200$. However the solution selected on the basis of BIC often corresponds to the case of slope parameters all set to zero, where it is worthless.

[Table 4 about here.]

Using the value of λ selected by cross-validation, the penalized estimates were compared with the maximum likelihood estimates (MLE) in terms of bias in absolute value and root mean square error (RMSE). Figure 4 shows the absolute bias of the penalized estimates against MLE, while Figure 5 shows the adaptive case. When the sample size is equal to 200, the bias of MLE is not negligible, while the penalized estimates present values substantially smaller in both the non-adaptive and adaptive cases. Non-adaptive penalized estimates are aligned with MLE when the sample size is equal to 500 and 1000, while the adaptive procedure tends to give a smaller bias. Non-adaptive penalized estimates are slightly more biased than MLE when the sample size is 5000. It is worth noting, however, that in this case the values are rather small. The bias of the adaptive penalized estimates is aligned with that of MLE for $n = 5000$. Overall, the adaptive version gives better results in terms of bias. Considering the RMSE, which is represented in Figures 6 and 7, penalized estimates present smaller values than MLE. The difference is important when the sample size is 200, and disappears as the sample size increases. In terms of RMSE, the non-adaptive penalization performs slightly better.

[Figure 4 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

[Figure 7 about here.]

5.2 Multidimensional case

The true item parameters taken for the simulation studies in the multidimensional case were set equal to the estimates obtained using the adaptive penalization in the second example, which are reported in Table 2. The sample size was set to $n = 200$, 500 and 1000. The latent variables were generated from a bivariate standard normal with independent components. The values of the tuning parameter λ were taken as in the unidimensional case. Both the non-adaptive and the adaptive penalization were applied, running only 50 replications for each setting due to the computational burden required by applying cross-validation for each dataset. Figures 8 and 9 show the absolute bias using the non-adaptive and the adaptive penalization respectively, while Figures 10 and 11 regard the RMSE. The figures show that both types of penalties reduce the bias, and that the non-adaptive penalization performs better especially in larger samples. The RMSE is definitely smaller with respect to MLE when applying the non-adaptive penalization. The adaptive penalization performs better than MLE, but not as good as the non-adaptive penalization. This is due to a few datasets where the MLE were extremely large, thus affecting also the estimates obtained with the small lasso penalty and hence the weights used for the adaptive penalization.

[Figure 8 about here.]

[Figure 9 about here.]

[Figure 10 about here.]

[Figure 11 about here.]

6 Conclusion

The nominal response model not only is suited for items with unordered response categories, but it is also useful for ordered response options because of its greater flexibility. However, this is achieved at the cost of the estimation of many parameters, which increases with the number of response categories. Hence, collapsing the response options is important to attain a more parsimonious model. This paper proposes a method to group the response categories based on a data-driven approach, exploring a lasso-type penalty and an adaptive version of it. The real-data examples and the simulation studies showed that the adaptive penalty combined with cross-validation is quite effective in grouping the response options.

Besides of grouping the categories, the method also yields regularized estimates of the parameters, and hence limits the variability of the estimates. In fact, penalized estimation yields a smaller root mean square error than maximum likelihood. The remarkable advantage of the penalized estimates is the reduction of the bias in small samples, with no relevant difference in large samples. Despite the penalty involves only the slope parameters, the intercept parameters benefit of the regularization as well.

Sometimes the ordering of the response categories is not respected because of the large variability of the estimates. This is probably the case of the first example given in this paper. The method proposed collapses these response categories indicating that the problem is not lack of ordering but that there is not enough information in the data for keeping the categories separate.

Finally, the method could also set all the slope coefficients of one dimension for one item to zero, if this is supported by the data, thus indicating that the latent variable is not related that item. Hence, the procedure is also able to perform variable selection and provide sparse solutions, which helps in interpreting the latent variables, especially in multidimensional models.

Thissen & Steinberg (1986) showed that the nominal model constitutes a family of models that includes other models, such as the partial credit and the 2PL model as special cases. Using a reparameterization of nominal model, they showed that other IRT models are obtained when some constraints are imposed on a design matrix. Using this approach, it is also possible to obtain models between the generalized partial credit model and the nominal model. Hence, in this approach the more constrained model is the generalized partial credit model, which is characterized by equally spaced slope parameters. The design matrix can also be constructed to constrain some parameters to be equal; however, this constitutes just an algebraic structure to express a class of models using a common formulation. In our approach, the slope parameters are forced to assume similar values, thus encouraging a fusion of the categories. The proposal does not hamper the interpretation of the results. If the slope parameters are grouped into two sets, for example, the procedure indicates to use a 2PL model for that item. In the other cases, the procedure leads to a nominal model, eventually with some response categories collapsed.

References

Arnold, T. B. & Tibshirani, R. J. (2014). *genlasso: Path algorithm for generalized lasso problems*. R package version 1.3.

Bartholomew, D. J., Steele, F., Galbraith, J., & Moustaki, I. (2008). *Analysis of multivariate social science data*. Boca Raton, FL: Chapman and Hall/CRC.

Bartolucci, F., Bacci, S., & Gnaldi, M. (2015). *Statistical analysis of questionnaires: A unified approach based on R and Stata*. Boca Raton, FL: Chapman and Hall/CRC.

Birnbaum, A. (1968). Some latent trait models. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, MA: Addison & Wesley.

Bock, D. R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37(1), 29–51. doi: 10.1007/BF02291411.

Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4), 443–459. doi: 10.1007/BF02293801.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1–29. doi: 10.18637/jss.v048.i06.

Eddelbuettel, D. & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8), 1–18. doi: 10.18637/jss.v040.i08.

Eddelbuettel, D. & Sanderson, C. (2014). RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis*, 71, 1054–1063. doi: 10.1016/j.csda.2013.02.005".

Fan, J. & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456), 1348–1360. doi: 10.1198/016214501753382273.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction.* New York, NY: Springer Verlag, second edition.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations.* Boca Raton, FL: Chapman and Hall/CRC.

Magis, D., Tuerlinckx, F., & Boeck, P. D. (2015). Detection of differential item functioning using the lasso approach. *Journal of Educational and Behavioral Statistics*, 40(2), 111–135. doi: 10.3102/1076998614559747.

Masarotto, G. & Varin, C. (2012). The ranking lasso and its application to sport tournaments. *The Annals of Applied Statistics*, 6(4), 1949–1970. doi: 10.1214/12-AOAS581.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16(2), 159–176. doi: 10.1177/014662169201600206.

R Core Team (2018). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Reckase, M. D. (2009). Multidimensional item response theory models. New York, NY: Springer Verlag.

Samejima, F. (1969). Estimation of ability using a response pattern of graded scores. *Psychometrika Monograph No. 17*. doi: 10.1007/BF0337.

Thissen, D. & Cai, L. (2016). Nominal categories models. In *Handbook of Item Response Theory, Volume One* (pp. 51–73). Boca Raton, FL: Chapman and Hall/CRC.

Thissen, D., Cai, L., & Bock, R. D. (2010). The nominal categories item response model. In M. L. Nering & R. Ostini (Eds.), *Handbook of Polytomous Item Response Theory Models*. New York, NY: Routledge. doi: 10.4324/9780203861264.ch3.

Thissen, D. & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, 51(4), 567–577. doi: 10.1007/BF02295596.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58(1), 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B*, 67(1), 91–108. doi: 10.1111/j.1467-9868.2005.00490.x.

Tutz, G. & Gertheiss, J. (2014). Rating scales as predictors—the old question of scale level and some answers. *Psychometrika*, 79(3), 357–376. doi: 10.1007/s11336-013-9343-3.

Tutz, G. & Gertheiss, J. (2016). Regularized regression for categorical data. *Statistical Modelling*, 16(3), 161–200. doi: 10.1177/1471082X16642560.

Tutz, G., Pößnecker, W., & Uhlmann, L. (2015). Variable selection in general multinomial logit models. *Computational Statistics & Data Analysis*, 82, 207–222. doi: 10.1016/j.csda.2014.09.009.

Tutz, G. & Schauberger, G. (2015). A penalty approach to differential item functioning in Rasch models. *Psychometrika*, 80(1), 21–43. doi: 10.1007/s11336-013-9377-6.

Wang, L., You, Y., & Lian, H. (2013). A simple and efficient algorithm for fused lasso signal approximator with convex loss function. *Computational Statistics*, 28(4), 1699–1714. doi: 10.1007/s00180-012-0373-6.

Wollack, J. A., Bolt, D. M., Cohen, A. S., & Lee, Y.-S. (2002). Recovery of item parameters in the nominal response model: A comparison of marginal maximum likelihood estimation and Markov chain Monte Carlo estimation. *Applied Psychological Measurement*, 26(3), 339–352. doi: 10.1177/0146621602026003007.

Yuan, M. & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B*, 68(1), 49–67. doi: 10.1111/j.1467-9868.2005.00532.x.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476), 1418–1429. doi: 10.1198/016214506000000735.

List of Figures

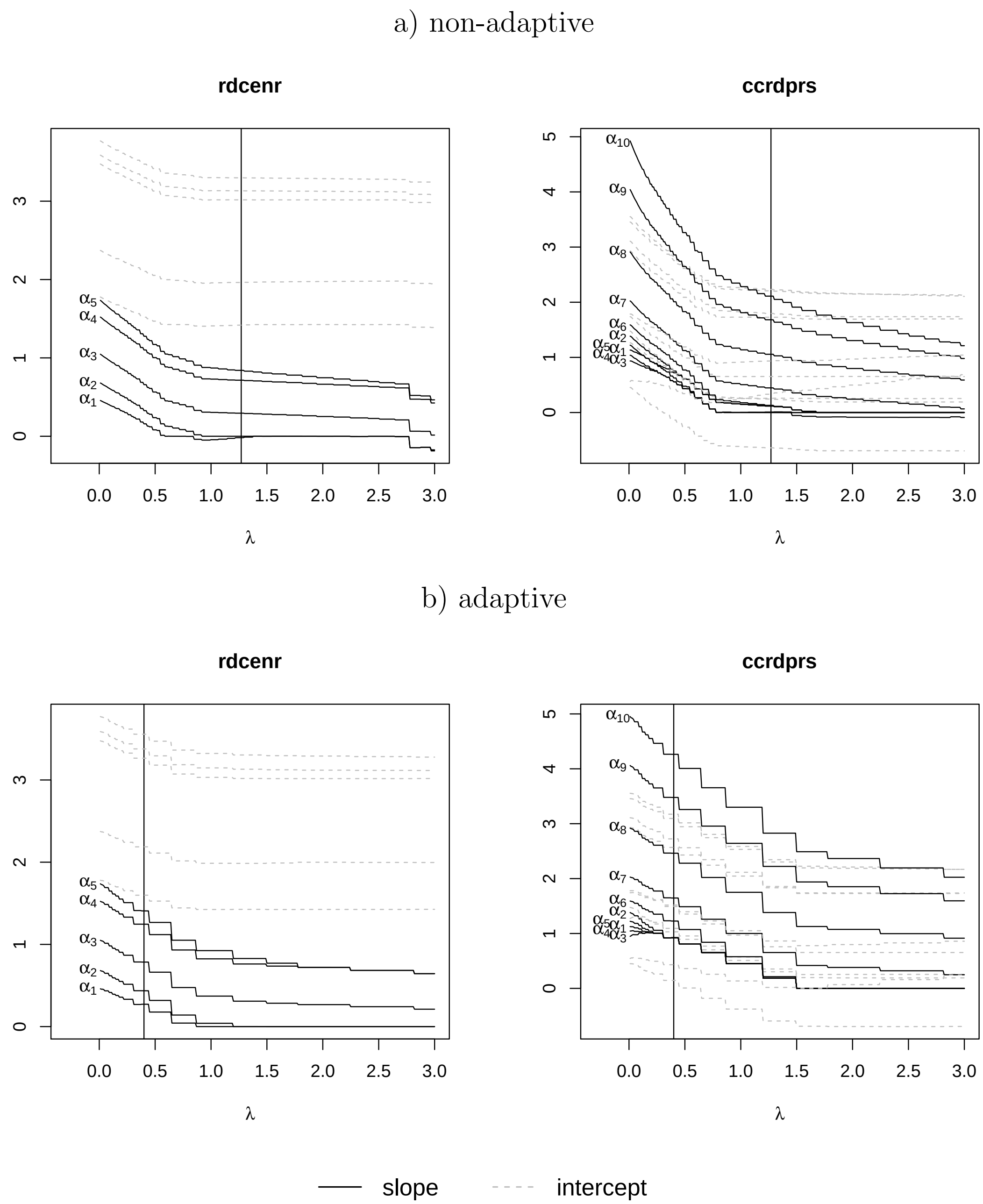


Figure 1: Regularization paths of two items of the European Social Study data.

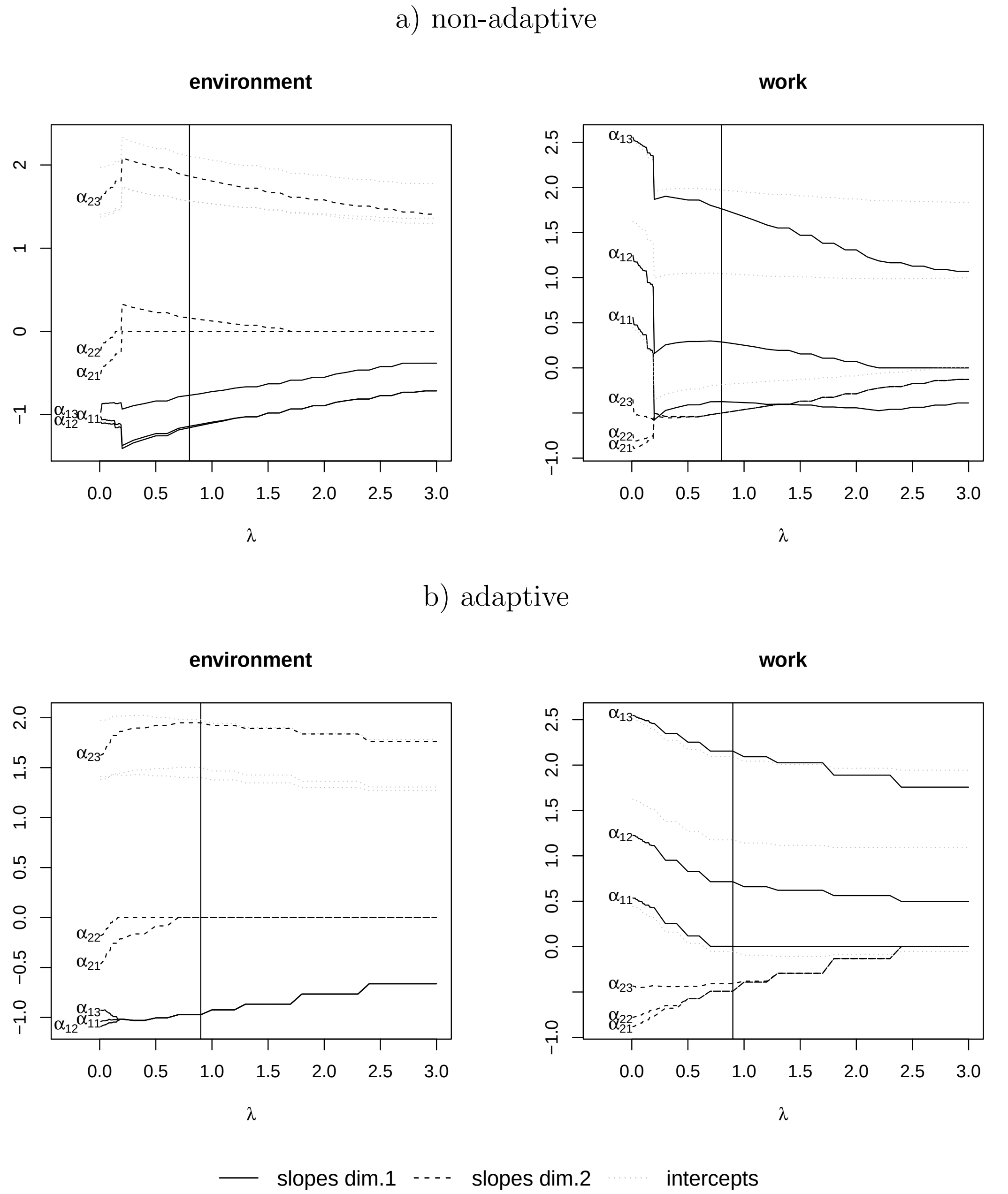


Figure 2: Regularization paths of two items of the Science and Technology data.

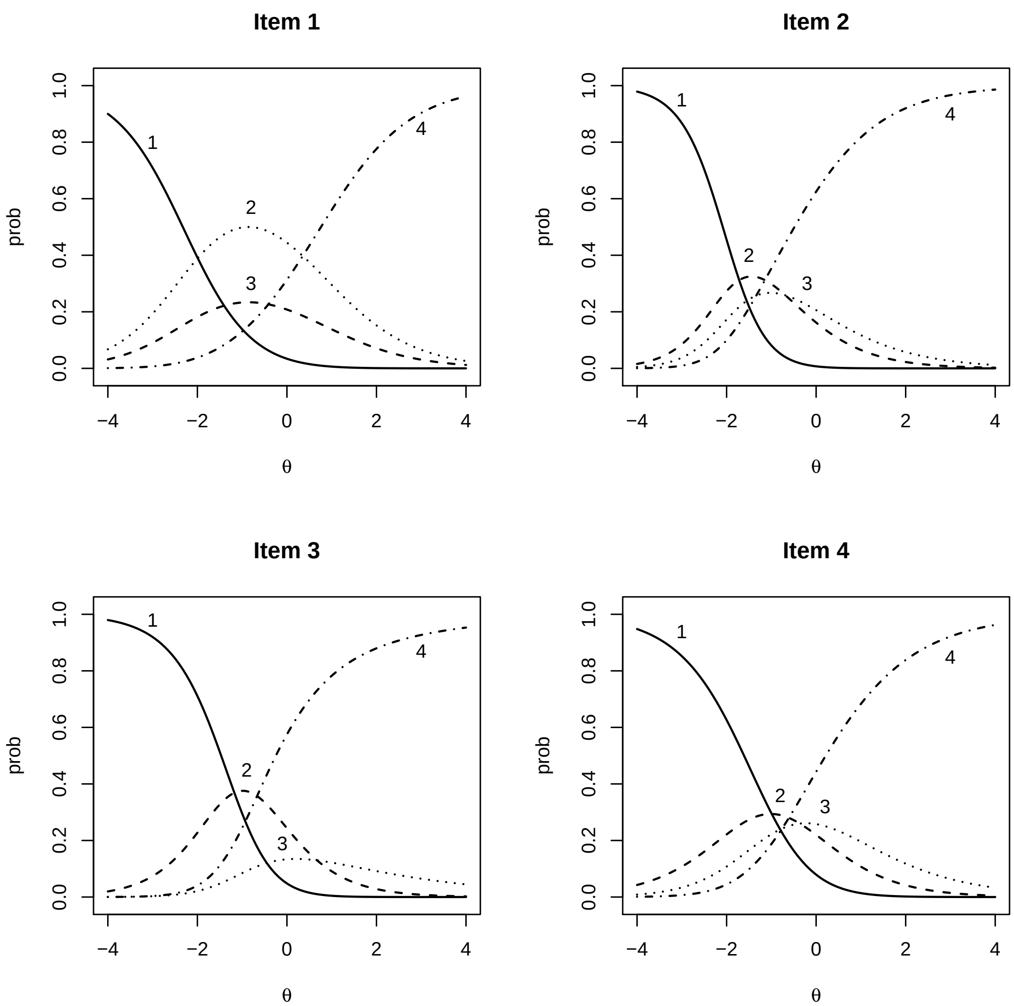


Figure 3: Probability curves of items used in the simulation study, unidimensional case.

Figure 4: Absolute bias of penalized estimates versus MLE, unidimensional case.

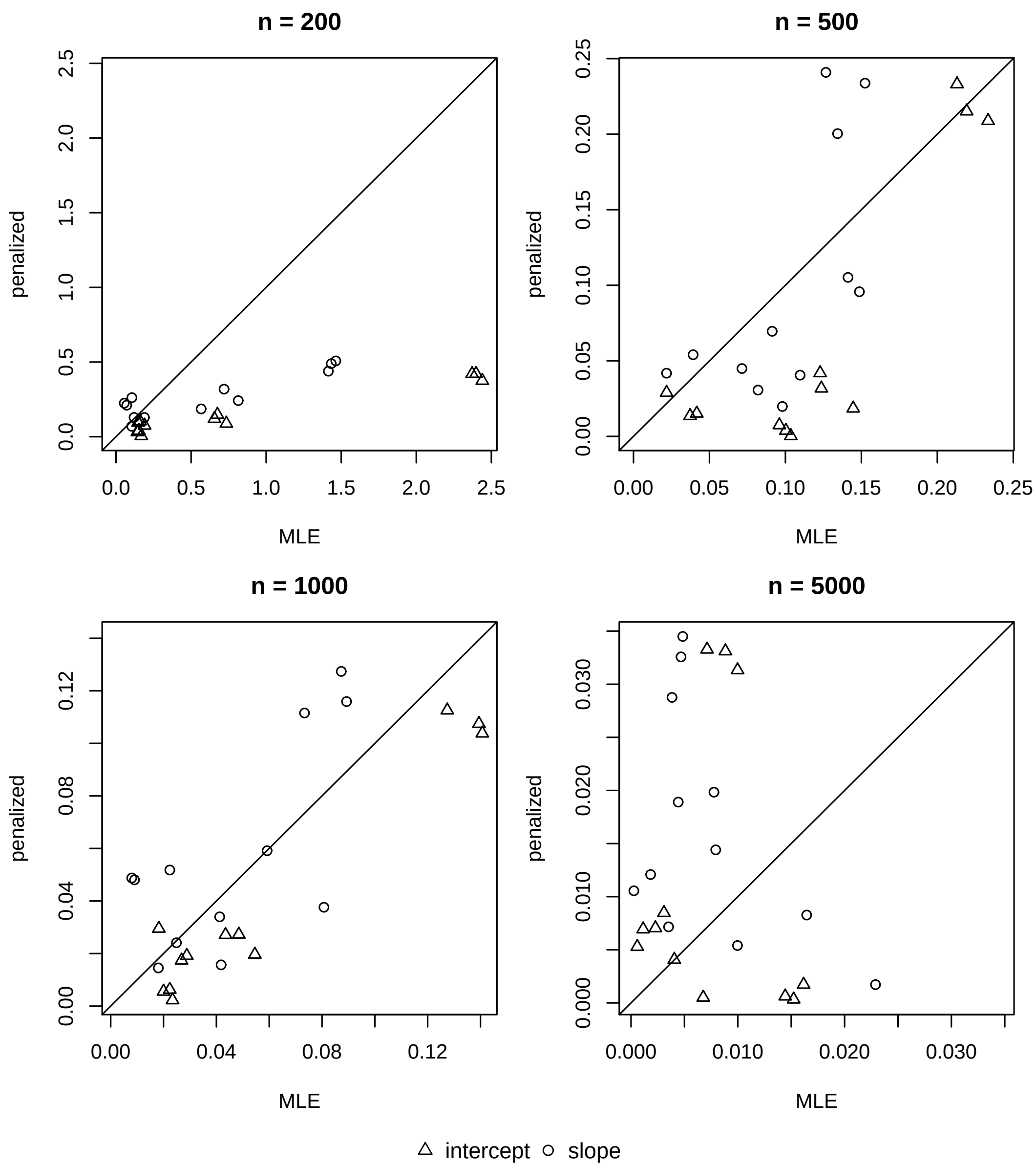


Figure 5: Absolute bias of adaptive penalized estimates versus MLE, unidimensional case.

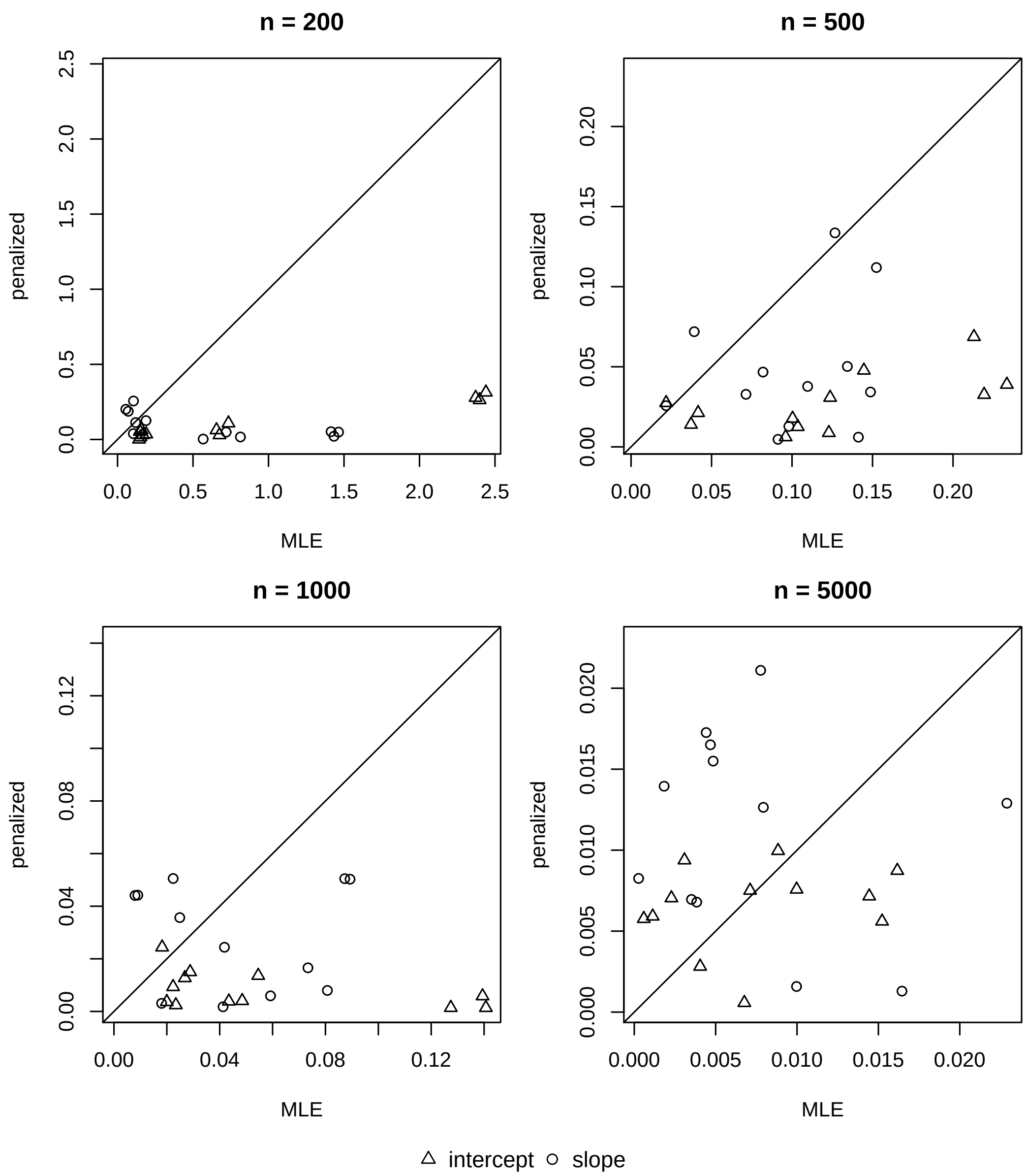


Figure 6: Root mean square error of penalized estimates versus MLE, unidimensional case.

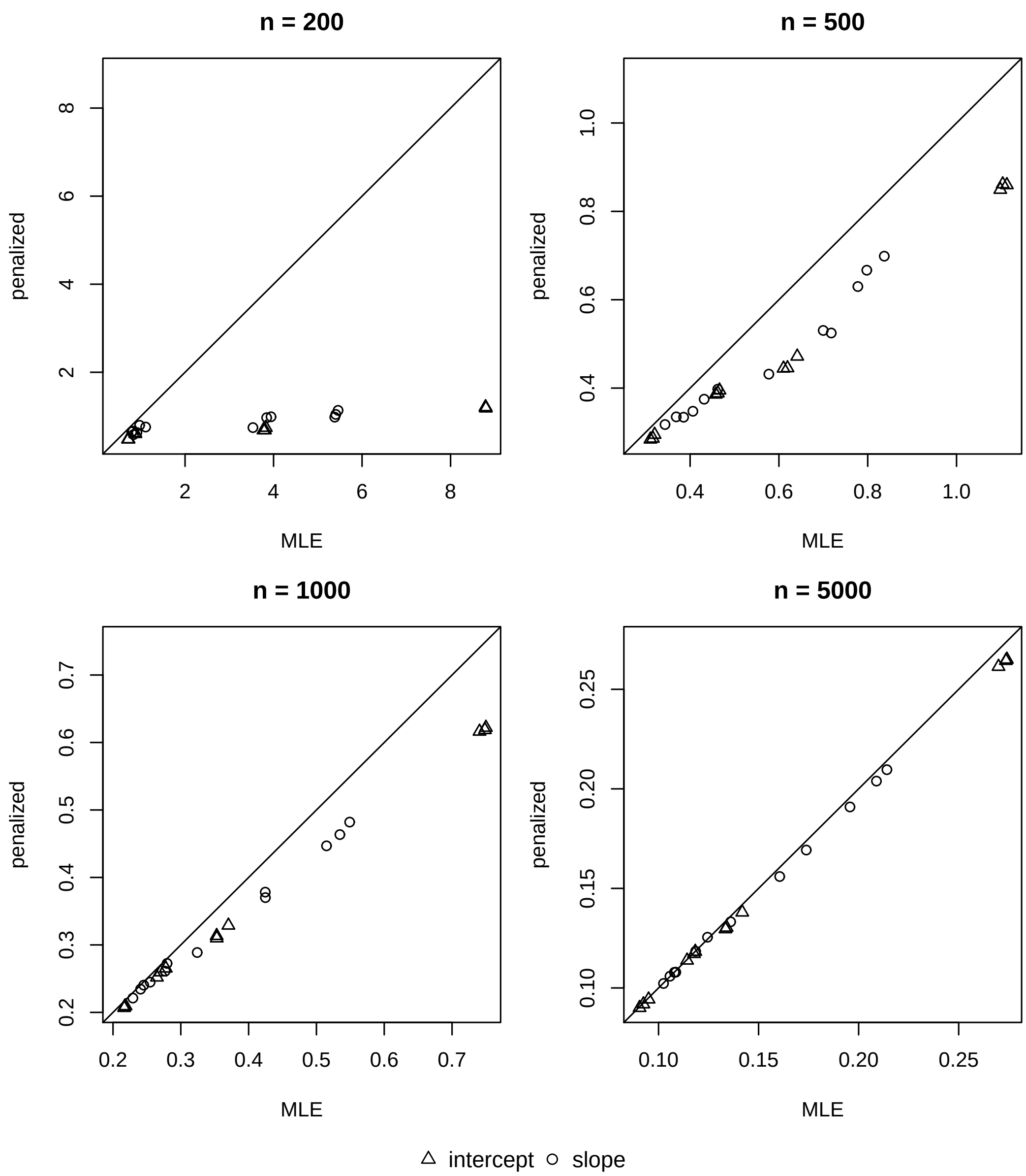


Figure 7: Root mean square error of adaptive penalized estimates versus MLE, unidimensional case.

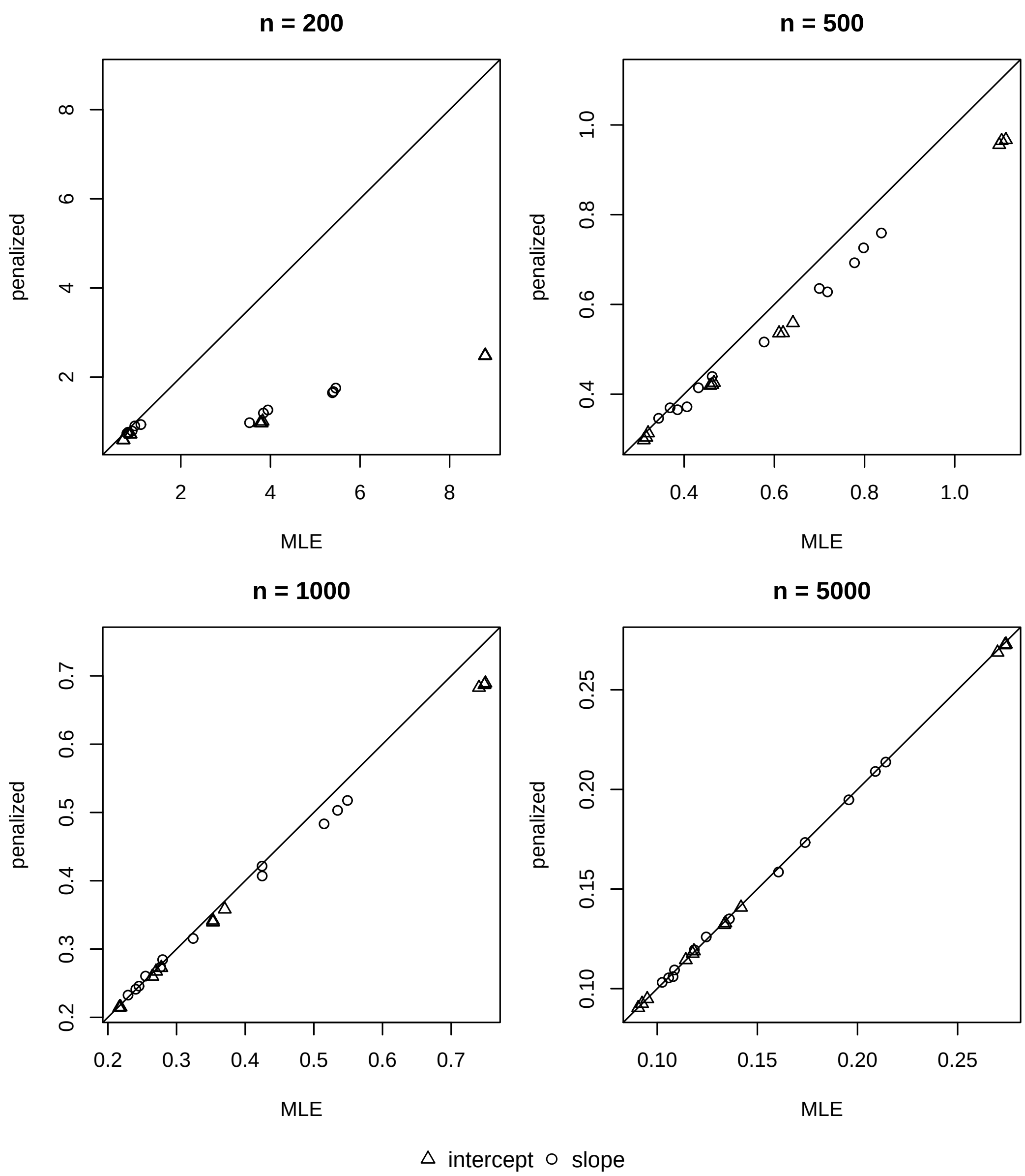


Figure 8: Absolute bias of penalized estimates versus MLE, multidimensional case.

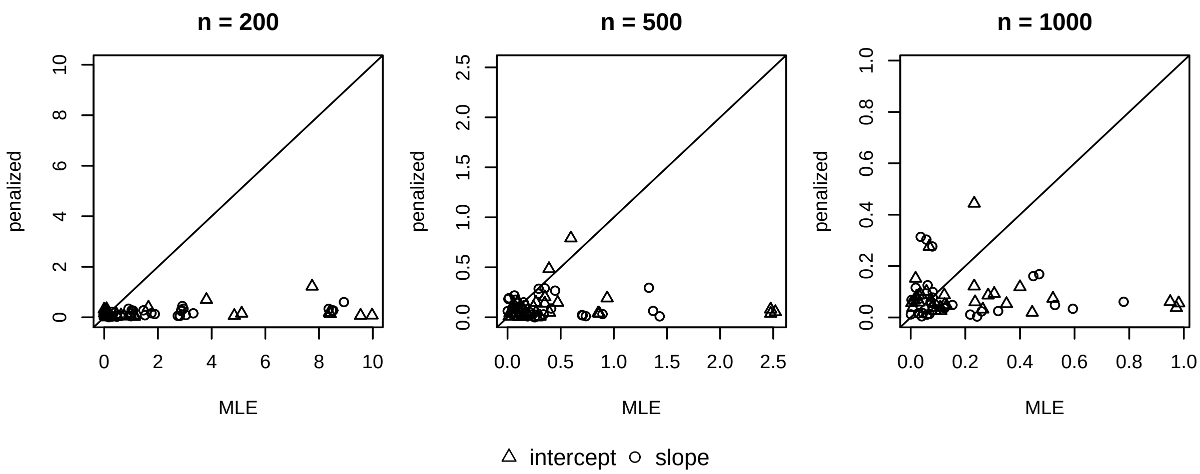


Figure 9: Absolute bias of adaptive penalized estimates versus MLE, multidimensional case.

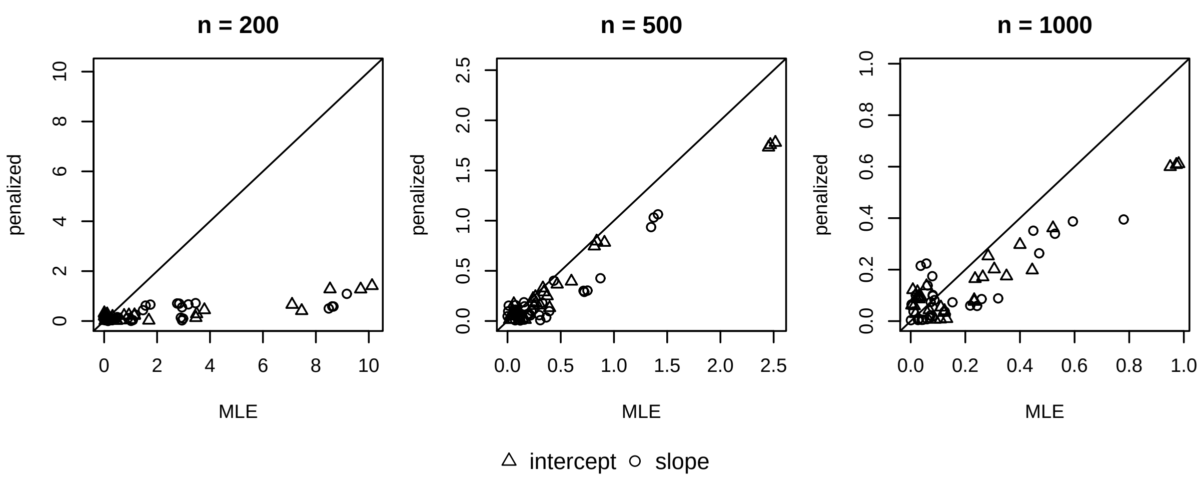


Figure 10: Root mean square error of penalized estimates versus MLE, multidimensional case.

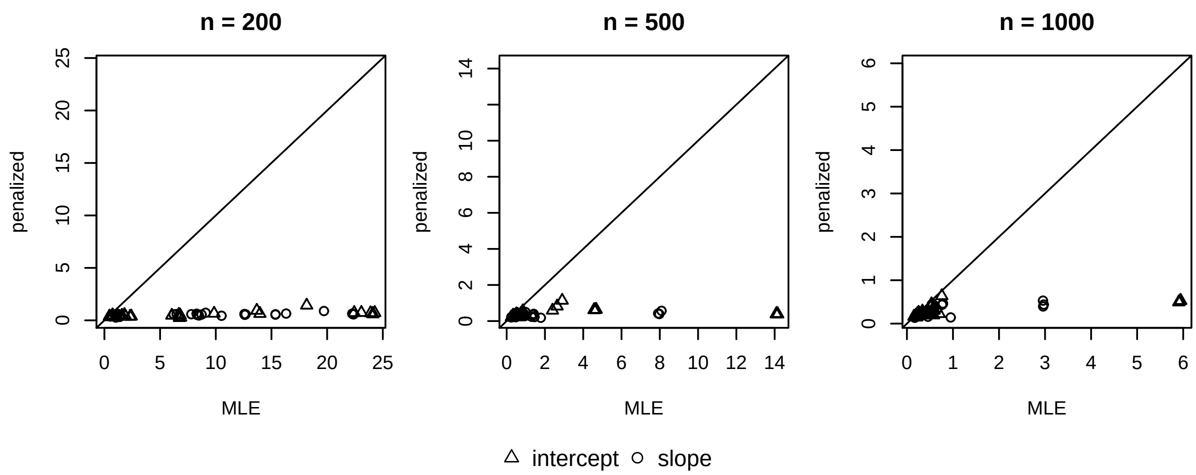
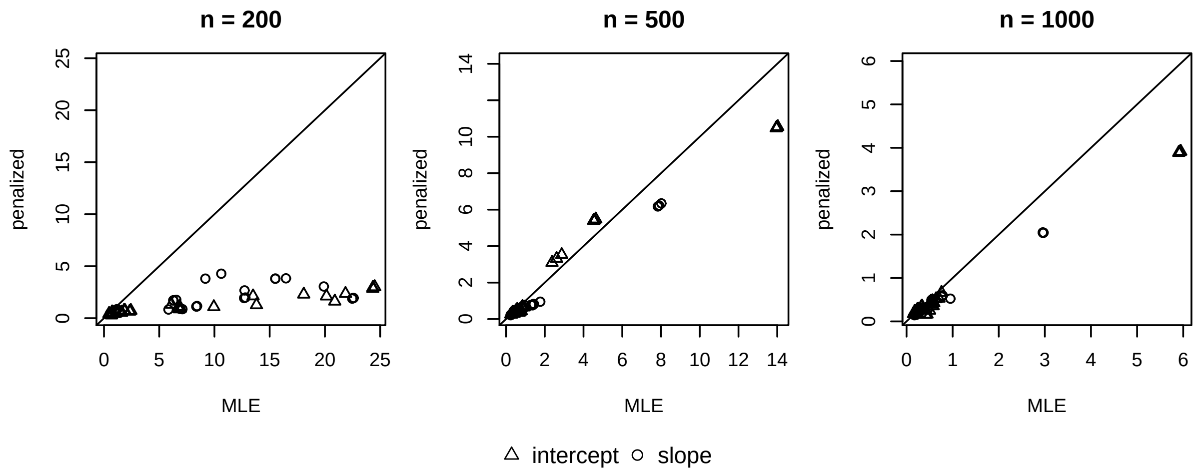


Figure 11: Root mean square error of adaptive penalized estimates versus MLE, multidimensional case.



List of Tables

Table 1: Slope parameter estimates at the selected value of λ, European Social Study data.

item	MLE	non-adaptive	adaptive
rdcenr	0.46	-0.02	0.27
	0.68	0.00	0.44
	1.05	0.29	0.79
	1.53	0.72	1.25
	1.74	0.84	1.41
ccrdprs	1.13	0.12	0.92
	1.39	0.12	0.92
	0.93	0.01	0.92
	1.04	0.00	0.92
	1.23	0.00	0.92
	1.59	0.13	1.23
	2.03	0.45	1.65
	2.93	1.06	2.46
	4.07	1.68	3.48
	4.97	2.11	4.26

Table 2: Parameter estimates at the selected value of λ obtained with the adaptive penalization, science and technology data.

item	α_{j1k}	α_{j2k}	β_{jk}
Comfort	0.35	-0.04	2.02
	0.50	-0.00	4.18
	1.45	0.21	2.83
Environment	-0.97	0.00	1.50
	-0.97	0.00	1.98
	-0.97	1.95	1.40
Work	0.00	-0.49	1.18
	0.72	-0.49	2.09
	2.15	-0.41	-0.05
Future	-0.01	-0.15	1.64
	1.35	-0.30	3.39
	3.60	0.04	1.53
Technology	-1.08	-0.41	1.93
	-1.08	0.00	2.67
	-1.08	2.10	1.83
Industry	-0.61	-0.00	1.63
	-0.61	-0.00	2.94
	0.00	2.43	2.55
Benefit	0.00	-0.37	1.60
	0.59	-0.37	2.41
	1.66	0.00	1.03

Table 3: True item parameters used in the simulation study, unidimensional case.

Item j	α_{j2}	α_{j3}	α_{j4}	β_{j2}	β_{j3}	β_{j4}
1	1.29	1.29	2.28	1.81	2.57	2.22
2	1.81	2.16	2.99	3.11	3.35	4.46
3	1.38	2.27	2.67	1.62	1.02	2.48
4	1.03	1.47	2.18	1.02	1.18	1.72

Table 4: Number of cases out of 500 in which α_{12} and α_{13} are fused at the selected value of λ, unidimensional case.

	n	200	500	1000	5000
cross-validation	non-adaptive	49	35	22	12
	adaptive	248	273	311	316
BIC	non-adaptive	366	3	2	3
	adaptive	324	83	61	40